

# Data Acquisition in Oceanographic Research Vessels

Oscar García, Joan Olivé, Dulce Afonso,  
Eduardo Arilla, Alberto Hernández, Juan Luis Ruiz,  
Xoan Romero, Antonio Sandoval,  
José Alberto Serrano, Jordi Sorribas  
Telematic and Communications Dept.  
Unidad de Tecnología Marina. CSIC. Barcelona, Spain  
e-mail: [utmtel@utm.csic.es](mailto:utmtel@utm.csic.es)

Enric Trullols, Joaquin del Río  
Applied Mathematics and Electronics Dept,  
Universitat Politècnica de Catalunya. BarcelonaTech  
Vilanova i la Geltrú, Spain  
e-mail: [enric.trullols@upc.edu](mailto:enric.trullols@upc.edu)  
e-mail: [joaquin.del.rio@upc.edu](mailto:joaquin.del.rio@upc.edu)

**Abstract** – This paper describes the design and the development of a new real time data acquisition system which focuses on oceanographic research. The system has been implemented in several Spanish oceanographic platforms (vessels and land-sites) using very reliable MESH network that can still operate when one node breaks down or a connection goes bad. Service Oriented Architecture allows expandable access to real time and historical data. This system is applicable to any network of instrumentation where an integrated management of data and a wide range of access procedures are needed. We can frame these data acquisition system in a satellite communications environment both VSAT (C band) and TNX75 (X band). The target domain off these projects is to provide long-range data and optimize satellite communications resources. The challenging issue is to achieve cross-platform data-generating infrastructure, which in an optimized use of system resources and provide real time information to the user without depending on the location in which to generate or produce them. The project is focused mainly in relation to the acquisition of data, such as an integrated system of generation of data with a multilayer structure for accessing data consulting services, all using communications via satellite. To this day, we fully integrated platform and running, and we are expanding the services and optimizing the end-user interfaces with these new services. The eyelet of this document is to provide global technical community the current status in Spain, in regard to oceanographic data generation and exploitation of the same by the scientific community, from both the oceanographic vessels located anywhere of the world from locations ashore.

*Keywords-data acquisition; web data services; OGC standards; service oriented architecture*

## I. INTRODUCTION

The Marine Technology Unit (UTM) from the Spanish Research Council (CSIC) manages three different oceanic research vessels: “Hesperides”, “Sarmiento de Gamboa” and “García del Cid”, and the Spanish Antarctic base “Juan Carlos-I”. These scientific platforms are devoted to acquiring

different kinds of scientific data, depending on the project (survey) in course. A new oceanographic data acquisition and exploitation system has been integrated on vessels managed by UTM. Several Data layers are implemented to allow data access from the more common tools and formats used in marine sciences. MESH networking was used in order to find a good solution for continuous connections and reconfiguration around broken or blocked paths. It also allows easy connection of new instruments and sensors.

Broadband communications systems via satellite installed on research platforms (ships and bases) provides a link from remote networks to the local area network of the UTM in Barcelona (Spain). A network of instrumentation on each platform works as a generator node of data, being all of them managed from the headquarters of UTM.

The objective is to ensure uniform treatment of information from these systems and their access (in real time or by consulting the historical records) by the maximum possible number of protocols and procedures.

It is designed following a Service Oriented Architecture (SOA) with a multilayer service schema in order to be easily expandable to new required data services [7]. It is fully operational and it has been implemented on top of the Java library and data model build in two previous technological research projects: The Virtual Distributed Lab for Oceanographic Data Processing and Monitoring (LabVir) and the Interoperability in Environmental and Marine Sensor Networks (INTIMAS), both funded by the Spanish government [6].

To achieve the centralization of data in a fast and stable environment, one of the problems that we face is to the dependencies of communications via satellite, and the optimized use of them, then detailing the steps we have taken for achieve good results.

The oceanographic data acquisition can be seen as a technological challenge in regard to getting the system becomes a real system for generating real-time data, and a resource server system for searching them all that using

current technologies in terms of infrastructure and communication protocols.

To this day, our communications infrastructure is implemented in oceanographic vessel BIO Hesperides and BIO Sarmiento de Gamboa. Regarding the vessel's communications Garcia del Cid, we are looking for alternatives of communication via satellite but with a limited geographical range to the Mediterranean Sea. On the other hand, the Spanish Antarctic Base Juan Carlos I, is implementing a VSAT (C Band) from this year and became fully operational during the next year (2011).

Regarding the data acquisition system, the system actually is fully operational, and its performance is like we expected, providing to our software architecture high values-added.

The implementation of the proposed data acquisition system, is conceived as a MESH network at the application layer, where all devices and instruments have the ability to talk, through web services, with the rest of equals with out any hierarchy.

In the next section we display the state of the art pointing the differences with previous works and showing the advantage of the MESH approach instead of the existing ones. Sections 3, 4, 5 and 6 are devoted to the communications infrastructure, the data acquisition core, the services offered by our architecture, and the data representation, respectively. Section 7 shows a case study working in one of our oceanographic research vessels. The paper ends with a conclusions and future work section.

## II. STATE OF THE ART

There are relatively few references to bibliographic data acquisition systems on oceanographic vessels and none with a level of integration similar to the proposal, detailing the services to low-level communications and acquisition (relations between instruments) to the upper layer interface user, where the scientific view and get the data produced. The integration of network management system with data management system is also new, although simple in concept, and there are no references to similar systems.

The MESH approach is advantageous in this area down as no predetermined relationships between systems that limit the expandability of the system (adding new instruments or sensors to the network) or to interfere with the restoration of the system against a potential failure of one of its elements. The latter concept is especially important in isolated environments such as neglected or acquisition systems that are installed on board ships or ocean observatories.

## III. COMMUNICATIONS

Broadband Satellite Communication Systems were installed in two vessels during 2008. In *"Sarmiento de Gamboa"* oceanographic vessel a VSAT Communication System in C band was installed with a data transfer rate of 192 kbps (CIR, Committed Information Rate) and a burst capacity

of 256 kbps (MIR, Maximum Information Rate) which can be increased to 5 Mbps. In *"Hesperides"* oceanographic vessel, following Navy specifications (the ship-owner) a TNX75 Communication System in X band was installed with a data transfer rate of 128 kbps (CIR) which can be also increased to 5 Mbps.

Apart from the advantages that an Internet access has had for the researchers, one of the most important characteristics of this Broadband System is the possibility of configuring a Virtual Private Network (VPN) between the vessels and the UTM main offices on shore (Land-Site). This network enables real-time data transmission and remote systems control and operation. Furthermore, web browsing is possible and an IP telephone system has been installed.

The broadband satellite service of UTM allows IP networking in permanent connection using VSAT technology.

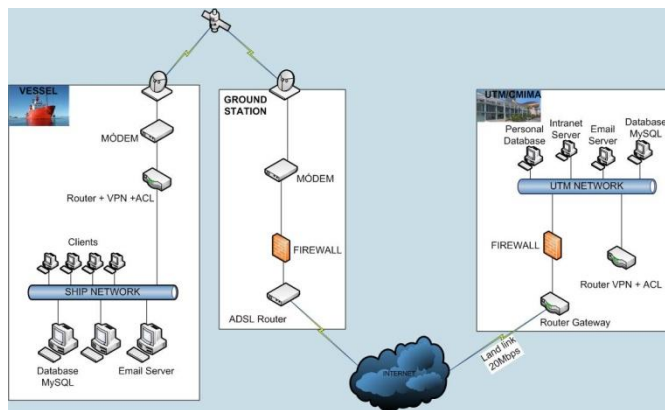


Figure 1. General schema of system communication on board *"Sarmiento de Gamboa"* and *"Hesperides"*

The satellite service providers are SEAMOBILE and HISDESAT. Both providers use geostationary satellites (orbiting at 35,786 km) and a large set of ground stations to provide coverage around the globe except the polar areas (their effective coverage is between 70 ° N and 70 ° S).

The terminals used onboard work with the C and X bands. The dimension of the antennas (2.4 m diameter) enables data transmission rates close to 5 Mbps in a global coverage scenario. With the current contract the providers guarantee a link with a minimum symmetrical bandwidth of 128Kbps. Since the key element of both systems is the VPN connection between the vessels and the UTM on shore (Land Site), the system architecture is virtually identical in both cases, although the receivers work in different bands. The diagram for that architecture is detailed in Figure 1, referring to the three locations involved: vessel, land backbone or teleport site, and UTM Land Site:

*Vessel Site:* On board there is a gyro-stabilized antenna and position control unit, a modem, and the network equipment (that enables IP routing, VPN establishment, access control lists, management of service quality, and links to the vessel network). In both vessels the IP equipment is the same brand and model (Cisco 2800 Series routers), while the electronic equipment for satellite link is specific to each case.

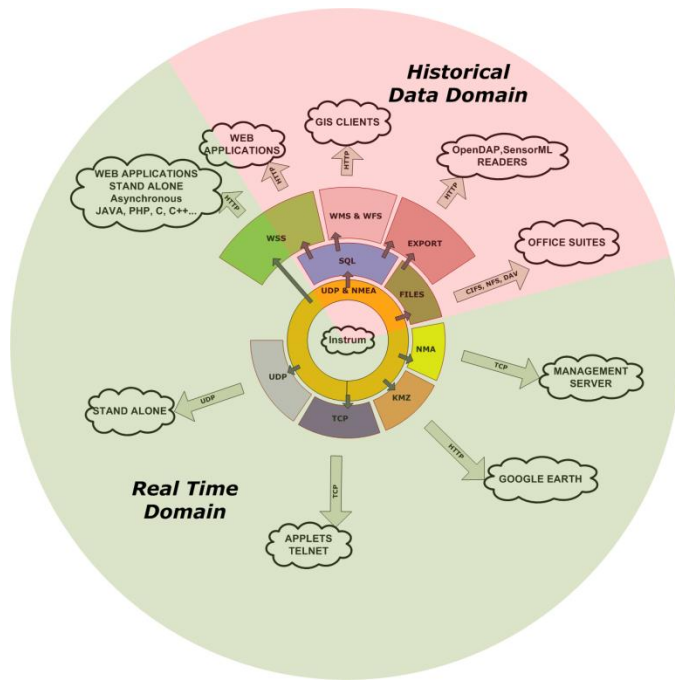


Figure 2. Data Service Layers, communication protocols used and data clients used in the Multilayer Service Data Acquisition and Operation System for Oceanographic ships and Instrumentation Networks. Real time and archive data domains are located at the background.

The configuration of these routers allows VPN establishment and Access Control Lists (ACLs) which allow specific computers to access Internet or Land Site Local Area Network.

**Teleport Site:** In order to link the satellites and the data backbone network (or public switched telephone network (PSTN)) providers need specific equipment. In this site a modem for the satellite link, an IP router for Internet connection and a firewall can be found. This last security element has been placed in this site in order to prevent the waste of bandwidth due to spam traffic.

**Land Site:** This equipment corresponds to the end of VPN and is located in the UTM Land Site (Barcelona, Spain). The VPN links are managed by a router. This is the central node in the Private Network which is extended between the UTM site and the oceanographic vessels. Only some computers have permission to access the ships' networks from the UTM LAN, these are specific in the ACL configuration.

#### IV. A MULTILAYER SERVICE DATA ACQUISITION AND OPERATION SYSTEM (SADO)

The basic mechanisms of transport and exchange of data between acquisition applications and interfaces, users or other external systems have been built using a model of overlapping layers (Figure 2). Basic interfaces, and instrumentation have been placed at the core of the system, and several layers of

service of increasing complexity are superimposed until the mechanisms of communication with user applications. Layers of service cover data access in real time as well as access to stored or archived data sets.

##### A. Basic Data Transport Service Layer

The basic mechanism for information exchange from acquisition applications to the different layers of service is made using datagrams ASCII following the NMEA-183 standard using the User Datagram Protocol protocol (UDP) in broadcast mode. The inner ring of the service layer model for communication with instruments is formed upon this basic mechanism that has been used and tested successfully from several previous works [2].

For instruments that use UDP as the native method of data transmission, this layer of service provides the proper normalization of telegrams that identifies the type of instrument and the platform (ship or base) that is issued. For those that use other input/output interfaces, an additional network interface is provided within this service.

The UDP protocol is not as reliable as TCP in the delivery of messages, but is faster, it also introduces low overhead of headers in the network and allows a design between client applications and servers without needing to establish permanent links or specific boot order when is used in broadcast mode.

##### B. Access to Historical Records Service Layers

In order to provide access to the historical records of the acquired data a three-layer service have been designed and developed.

In each platform and in the central node of the whole system the historic records are feed by real time acquisition applications while an additional copy is stored, every minute, in the central node by database synchronization procedures. This databases redundancy allows the detection of sporadic cuts in satellite links.

- 1) *RAW and Processed Files Service Layer (FILE).* This layer provides data storage in ASCII files with the raw information, contained in the telegrams NMEA and ASCII files in CSV format [9], with a subset of data encoded within the telegrams specifying the name and units of the measured variables. The ASCII files are generated daily. The name of each file is a code with information about the day, month and year of its creation..The extension of the name identifies the instrument that originates the data.
- 2) *Relational Database Service (SQL)* This layer provides data storage in tables in a relational database with geographic extensions, as POSTGRES with PostGIS. The acquisition date adjusted to the millisecond is formed as primary key databases. The relationships between tables are not formally

expressed as table relations, but are implemented inside the logic of the service at the application level.

- 3) *Geographic Data Service (WFS and WMS)* It is built on the previous layer. This service offers georeferenced data in map and features using the standard WMS and WFS from Open Geospatial Consortium [8]. The requests to these two services are initiated by GIS clients using the HTTP protocol as a channel of communication with the servers WMS / WFS. This is a service typically oriented to recover historical data but its design allows its operation in real time by defining an specific "service" for taking the last data arrived in the system.
- 4) *Data Export Service (EXP)* An exportation interface to different formats commonly used in marine and environmental sciences as SensorML [13], NetCDF [9] or openDAP [10] is built from the two previous layers.

### C. Real Time Data Access Service Layers

These layers are formed by all the service layers that provides access to data in the same moment of its acquisition, in different formats and network protocols, asynchronously or by request.

- 1) *UDP Data Service* This service multiplex UDP messages received on an specific port to multiple ports in broadcast mode or directed to a particular IP address. This service is oriented to processes or applications such as real time data graphic displays or event handlers.
- 2) *TCP Data Service.* This service replicates the messages received on a specific UDP port to three TCP ports in three different formats: ASCII raw from the received NMEA, self describing XML with information of structure of corresponding data and Java object as a serialized object of data from the used LabVir library. It is a service oriented to other processes or applications that need to overcome the inherent limitations in the use of UDP, such as the multi attendance to a determined port number in the same server, the reliability of the data reception, and the overcoming security mechanisms on broadcast traffic related to Java applications. This layer provides an interface to universal queries of on-line data, using a valid TCP client that is a tool from all the operating systems that support TCP/IP as "telnet". TCP data service implements a mechanism to match the UDP telegram reception rate to its asynchronous service to TCP clients.
- 3) *WSS Data Service.* This service offers real-time data provided by UDP telegrams and the stored data from

databases in Service Oriented Access Protocol (SOAP) using a Web Service called Web SADO Server (WSS) that are published on a application server using WSDL [6,11] for its description. The data access using WSS enables communication from the data acquisition system to any application, regardless of the programming language or platform used. Part of WSS service is built on the SQL service layer becoming a layer from access to historic data layer, and part is built over the Basic Data Transport Service Layer in order to provide data access in Real Time.

- 4) *KMZ Data Service.* This service provides the values of the last acquired data in compressed KML files [14] whose content and structure is updated in real time and are served through HTTP protocol. In those files a time value is included in order to allow the automatically refreshment of data. This is a service oriented to Google Earth clients, although some map servers and Geographical Information System clients can define contents layers from those files.
- 5) *Data Management Agent Service (NMA).* This layer allows the asynchronous transmission of data using a TCP channel to a Zabbix network management [12] server. Through this service, the functionality of the management system used to monitor the physical network devices is extended by introducing the quality of data management.

## V. UTM GIS ARCHITECTURE

The geographic information system developed by the UTM have as a principal goal to give facilities to access and work with the data that collects the research oceanographic vessels "Hespérides", "Sarmiento de Gamboa", "García del Cid" and the Spanish Antarctic base "Juan Carlos I" in Livingston island. This system is easily scalable to accept more platforms and data sources like automatic sub-aquatic stations, buoys and new vessels. All this information has an important spacial and temporal component.

The UTM GIS solution set it's all developed over Open Source projects and take into account the OGC (Open Geospatial Consortium) standards. That grants the interoperability of the data and the services over the data. We have identified and developed four principal components:

*Georepository:* store the information of the system and guarantees the persistence of the data. It stores all the spatial and thematic information or a reference to the thematic information. It uses PostgreSQL as a relational data base with the spatial extension POSTGIS.

*Geographic information server:* it uses Geoserver and offers publication map services: WMS, WFS and WCS over the OGC standards. The geographic information server has been deployed over the application server GlassFish. Geoserver is



capable to integrate heterogeneous data sources like shape files, PostGIS and MySQL databases , KMZ, GoogleMaps, GML and more. Geoserver can do Tiling over the visualization layers: cache preload of the cells around the one that we are seeing now.

*Web client:* the web client is developed with the Javascript Framework: OpenLayers using the WMS and WFS services from the GeoServer. With this, it's possible to offer an easy data access and a simple 'first tool' to work with this data: it's possible to calculate areas, distances, selection and edition of different data layers and pdf printing.

*Thin client:* Desktop multiplatform GIS application java developed with custom plugging over the Kosmo Open Source project. This client has the most common functionality for this environment. Raster and Vector in their different formats that the system stores into ordered directories structure in the server.

The information is stored in the relational data base PostgresSQL with the PostGIS extension that gives support to geographic objects. PostGIS stores the geographic information in a Geometry data type column.

All the data is saved in a central server. There are two main ways to store the data depending on the data source and the use of the data:

The data that is stored in the Georepository it's served with OGC Web Services. OGS services are the integration of

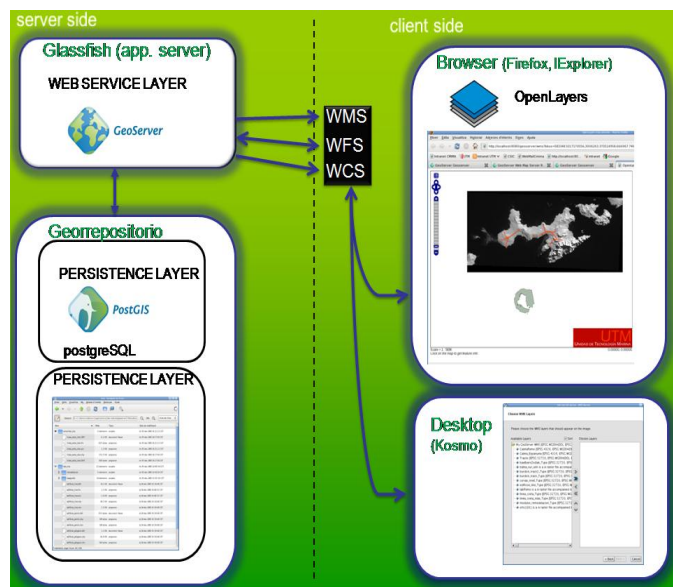


Figure 3. UTM GIS Architecture with the different components in the server side and thin web client and desktop GIS software accessing the data in the client side.

different OGC specifications, focused on geoprocessing (WMS, WFS, WCS, WTS, etc.), using XML and HTTP technology.

The geographic information server, Geoserver, offers these services over the data:

*WMS:* Geoserver serves this georeferenced formats: jpeg, gif, png, svg, pdf, kml, kmz, that are generated dynamically. It's accessible using a web browser throw standard URL (Uniform Resource Locators) requests.

*WFS:* equal to WMS but it's possible to interact with the server maps. The interaction is with GML, derivation from XML.

*WCS:* geospatial data in coverage format. As a difference with WMS the WCS service provides the data in original semantic making possible to work directly with it and not only with the static representation of it, like WMS.

## VI. DATA REPRESENTATION

The UTM GIS Data Architecture offers different ways to easily access the information depending on the way that we are going to use it: real time data, historical data and different representations. All the applications are based and developed over Open Source software and are Open Geographic Consistorium (OGC) complaints.

*Map services:* WMS, WFS and WCS make possible an easy access to all the information in two ways: Directly with a web browser (Firefox) or from a GIS desktop application like Kosmo, ESRI, QGIS or gvSIG for example. Throw UTM GIS Web application it's possible to load different layers and work with them measuring distances, areas or adding points for example. This application is developed using Open Layers framework.

*OGC Sensor Observation Service (SOS):* UTM SOS offers SensorML data representation according to OGC standards for real time data acquisition of instrument and data description using OOSTethys Server, over two of our Oceanographic Research Vessels: Sarmiento de Gamboa and Hespérides. The UTM SOS Service is implemented in Perl and it consists in a CGI that offers three services: GetCapabilities, DescribeSensor and GetObservation.

*Real Time Data Viewer (RDV):* to visualize data in real time in different output formats: tabular, graphic or image. Data generated in the sensors are processed for the Data Turbine Server that is capable to do Data Streaming in ASCII and binary formats. Data Turbine is the data source for the RDV.

*Historical Data Viewer:* we use Zabbix [12] to monitor and visualize historical data. It's possible to visualize graphics and configure alarms to control data acquisition events for one data event or to combine different events to shoot an alarm. When an alarm is triggered it is possible to make different

actions like insert a new register to control it in a data base or to send an email for example.

## VII. CASE STUDY

As a particular case of the implementation, we describe the system that is working right now in our oceanographic research vessel “Sarmiento de Gamboa”.

The communications system implements a broadband Very Small Aperture Terminal working in C-band, a router on board the ship, and another router in the land site data center in Barcelona. Apart from the advantages that an internet access has had for the researchers, one of the most important characteristics of the system is the possibility of configuring a Virtual Private Network and the possibility of exchange real time data between the ship and land site in Barcelona.

As data source we have a server, which is responsible for collecting the data traveling over the network (UDP frames, NMEA datagrams), store this data in our databases (MySQL and Postgres in real time), and also integrate them into data files in both raw and processed form.

A geographic service layer georeferences our data using WMS WFS standards.

To access data in real time, we have established multilayer services, which we offer the following features, UDP frame generation, replication of data with TCP frames, web services access to data in databases, file generation KML, asynchronous transmission of data with the Zabbix server.

With these data, and by other agents, we provide information to Zabbix and DataTurbine systems, which offer us the service of access to data in real time, as historical trends querying.

## VIII. CONCLUSIONS AND FUTURE WORK

A new data acquisition system has been described. The system has been implemented in several Spanish oceanographic platforms. The system is applicable to any network of instrumentation where an integrated management of data and a wide range of access procedures are needed.

Service Oriented Architecture allows expandable access to real time and historical data. In the future, new instruments will be added to the current integrated data network, especially those which have complex data formats or those which have a high data streaming behavior. Data access will be improved by the creation of a new metadata service layer.

In the future the system will be implemented in new platforms as cabled seafloor observatories or remote weather stations located in Antarctica.

## ACKNOWLEDGMENT

This work has been supported by the Spanish “Ministerio de Ciencia e Innovación” under contract CTM2009-08867.

## REFERENCES

- [1] T.R. Henderson and R.H. Katz, “Transport Protocols for Internet-Compatible Satellite Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 17, No. 2, pp. 345-359, February 1999.
- [2] M. Allman, D. Glover, and L. Sanchez, “Enhancing TCP Over Satellite Channels using Standard”, 1998.
- [3] A. Tirumala, M. Gates, F. Qin, J. Dugan, and J. Ferguson, “The TCP/UDP bandwidth measurement tool”. <http://www.nlanr.net/>. Last visit 19/05/2010.
- [4] H. Newton, “Newtons Telecom Dictionary”, CMP Books 2001.
- [5] W. Stallings, “Data and computer communications” Fifth Edition, Prentice Hall International, 1997.
- [6] E. Trullols, J. Sorribas, J. del Río, C. Samitier, A. Manuel, and R. Palomera, “LabVir: A Virtual Distributed Measurement System,” *IEEE Instrumentation and Measurement Technology Conference*, Colorado, USA, May 2003.
- [7] J. Sorribas, A. Tudela, A. Castellón, O. Chic, Z. García, J. Prades, and D. Montero, “An oceanographic data acquisition system for Ethernet LAN for Spanish Research vessels”, *IEEE Oceans’98 Conference Proceedings*, vol. 1, San Diego. USA, pp. 93–97, 1998.
- [8] Open Geospatial Consortium Standards.
- [9] <http://www.opengeospatial.org/standards/common>. Last visit 19/05/2010
- [10] Network Common Data Format.  
<http://www.unidata.ucar.edu/netcdf>. Last visit 19/05/2010.
- [11] Open-source Project for a Network Data Access Protocol.
- [12] <http://opendap.org/>. Last visit 19/05/2010.
- [13] Web Service Definition Language. <http://www.w3.org/TR/wsdl>. Last visit 19/05/2010.
- [14] “An enterprise-class open source distributed monitoring solution”. <http://www.zabbix.com/>. Last visit 19/05/2010.
- [15] SensorML. <http://marinemetadata.org/references/sensorml>. Last visit 19/05/2010.
- [16] Google Earth, KML.  
[http://earth.google.es/userguide/v4/ug\\_kml.html](http://earth.google.es/userguide/v4/ug_kml.html). Last visit 19/05/2010.